

SCHEDA DI CASO

RISULTATO ATTESO 1 - Eseguire operazioni di raccolta dati e data mining, provvedendo alla preparazione per i successivi trattamenti

1 - RACCOLTA E PREPARAZIONE DATI

Grado di complessità 4

1.4 DATA PREPARATION

Promuovere operazioni di data cleaning, eliminando possibili errori e stabilendo dei meccanismi di comportamento in caso di dati mancanti, e di pre-processing, esplorando il target data ed effettuando un'analisi su base campionaria

Grado di complessità 3

1.3 SELEZIONE DI UN TARGET DATA SET

Selezionare, decodificare e segmentare i dati grezzi secondo criteri prestabiliti al fine di pervenire ad un sottoinsieme di variabili e di dati o di un campione di dati (structured data, unstructured data) che rappresentano un determinato target data o dati obiettivo

Grado di complessità 2

1.2 TRANSFORMATION & CONFIGURATION

Effettuare operazioni di trasformazione, convertendo tipi di dati in altri o definendo nuovi dati ottenuti attraverso l'uso di operazioni matematiche o logiche sulle variabili, e configuration, effettuando una loro riconfigurazione in caso di provenienza da fonti diverse

Grado di complessità 1

1.1 DATA GATHERING

Raccogliere dati e informazioni "grezze", disponibili in diversi format (.pdf, .csv, .txt, ecc.), utilizzando fonti attendibili, certificate e coerenti con il fabbisogno richiesto

2 - OPERAZIONI DI DATA MINING

Grado di complessità 2

2.2 SCELTA DEL PROCESSO DI DATA MINING

Stabilire la tipologia di tecnica di data mining da utilizzare (cluster analysis, regression analysis,

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

classification analysis, anomaly detection analysis, intrusion detection, association rule learning, decision tree, neural networks, rule induction, data warehouse) più idonea rispetto all'obiettivo dell'analisi

2.2 SCELTA DEGLI ALGORITMI DI DATA MINING

Selezionare il metodo da usare per ricercare patterns nei dati, definendo quali parametri possono essere più appropriati e integrando i metodi di data mining scelti con i criteri generali di Data Analytics (Knowledge Discovery in Databases)

Grado di complessità 1

2.1 DATA MINING

Adattare l'algoritmo standard prodotto al caso preso in esame, in base alla tipologia di data mining selezionata (cluster analysis, regression analysis, classification analysis, ecc.), applicandolo ad relativo database e fornendo come outcome la descrizione dell'informazione ricercata

3 - CONSOLIDAMENTO DATI

Grado di complessità 1

3.1 CONSOLIDAMENTO DEI MODELLI IDENTIFICATI

Analizzare e verificare i risultati ottenuti e le performance del sistema, anche valutando possibile retroazione a fasi precedenti, al fine di migliorare l'efficacia dei modelli di data analytics

SCHEDA RISORSE A SUPPORTO DELLA VALUTAZIONE DEL RISULTATO ATTESO 1

RISORSE FISICHE ED INFORMATIVE TIPICHE (IN INPUT E/O PROCESS ALLE ATTIVITÀ)

- Obiettivi di analisi
- Norme in materia di protezione della privacy e protezione dati personali (GDPR)
- Criteri generali di Data Analytics (Knowledge Discovery in Databases)
- Data Analytics (Alteryx, Apache Spark, SAS, Big ML, MATLAB, Jupyter)
- Structured data
- Unstructured data
- Dati grezzi (.pdf, .csv, .txt)

TECNICHE TIPICHE DI REALIZZAZIONE/CONDUZIONE DELLE ATTIVITÀ

- Tecniche di data cleaning
- Tecniche di data processing
- Tecniche di data transformation
- Tecniche di data configuration
- Tecniche di data gathering
- Cluster analysis
- Regression analysis
- Classification analysis
- Anomaly detection analysis
- Intrusion detection analysis
- Association rule learning
- Decision tree
- Neural networks
- Rule induction
- Data warehouse

OUTPUT TIPICI DELLE ATTIVITÀ

- Raccolta e preparazione dati effettuata (data cleaning, data transformation e configuration, data catering)
- Algoritmo di Data Mining selezionato e adattato
- Descrizione dell'informazione ricercata

INDICAZIONI A SUPPORTO DELLA SCELTA DEL METODO VALUTATIVO E DELLA PREDISPOSIZIONE DELLE PROVE

ESTENSIONE SUGGERITA DI VARIETÀ PRESTAZIONALE

1. L'insieme delle tecniche e metodologie di data gathering e preparation

ADA.14.01.20 - DATA SCIENCE AND ANALYTICS

2. L'insieme delle tecniche e metodologie di data transformation e configuration
3. L'insieme di regole relative alla data mining
4. Un set di obiettivi di analisi
5. Un set di dati grezzi

DISEGNO TIPO DELLA VALUTAZIONE

1. Prova prestazionale: a partire dai set dati, simulazione delle operazioni di raccolta e preparazione dati (data gathering, data preparation, data cleaning, data transformation & configuration) e data mining (con scelta di una delle tecniche a disposizione)
2. Colloquio tecnico relativo alle scelte adottate nella raccolta e preparazione dei dati e nell'applicazione dell'algoritmo standard per le operazioni di data mining

FONTI

Cesarini, M., Fontana, M. Mercurio, F., Mezzanecanica M. & Vegetti, N., Data Quality: Un Approccio Metodologico e Applicativo. “Il caso delle COB del mercato del lavoro in Lombardia”, Working Paper “Documentazione relativa alla ricostruzione di una metodologia unificata, ripetibile e aperta, CRISP E ARIFL EMC (2015), Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, John Wiley & Sons, Indianapolis
European e-Competence Framework (E-CF), v.04: <https://www.ecompetences.eu/>
Freitas, A. & Curry E. (2016). Big Data Curation, In Cavanillas, J.M & Curry E. (2016). New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe, Big Data Public Private Forum, Springer Open, p. 87-119
Gelernter, J. & Lesk, M. (2011), Use of Ontologies for Data Integration and Curation. In The International Journal of Digital Curation, Issue 1, Volume 6

SITOGRAFIA

MediaSapens, Data Curation: the essential step for integrated data-driven research